Received: 11 October 2024 | Accepted: 13 July 2025

DOI: 10.1111/2041-210X.70118

Methods in Ecology and Evolution

RESEARCH ARTICLE

AniDomNet: A sequential pairwise model for inferring dynamic animal dominance hierarchies

Nusret Ipek[1] | Frank A. M. Tuyttens[2,3] | Bernard De Baets[1] | Jan Verwaeren[1]

[1]Department of Data Analysis and Mathematical Modelling, Ghent University, Ghent, Belgium

[2]Animal Sciences Unit, ILVO, Merelbeke-Melle, Belgium

[3]Department of Veterinary and Biosciences, Faculty of Veterinary Medicine, Ghent University, Merelbeke-Melle, Belgium

Correspondence
Nusret Ipek
Email: nusret.ipek@ugent.be

Funding information
Bijzonder Onderzoeksfonds UGent, Grant/Award Number: 01D01522

Handling Editor: Andrea Ravignani

Abstract

1. Inferring dominance hierarchies is key to quantifying social dynamics within animal groups. Observed dyadic agonistic interactions remain an important source of data for studying dominance hierarchies. As a result, numerous (statistical) approaches attempt to derive and characterize dominance hierarchies from dyadic interactions. However, most of them ignore the temporal component of these interactions.
2. We introduce a novel model to characterize dominance hierarchies using a sequential pairwise relationship model called Animal Dominance Network (AniDomNet). This model is inspired by the Elo ranking model, yet relaxes several of the underlying assumptions and allows us to study the dynamics of hierarchy formation.
3. While addressing certain shortcomings of the current sequential methods, AniDomNet also excels at predicting the outcome of future interactions. Moreover, we propose a social agony-based approach to obtain a directed acyclic graph (DAG) that represents the dominance hierarchy according to a fitted model.
4. AniDomNet is shown to be a useful tool to detect mistakes (such as identity switches) made during the observation process.

KEYWORDS
animal behaviour, dominance hierarchy, Elo rating, modelling, sequential approach, sociometry

1 | INTRODUCTION

Across diverse taxa of gregarious animals, unravelling dominance relationships between individuals is key to understanding the social structure in animal societies. Dominance relationships between animals, and at a higher level the emergent dominance hierarchy, can have an impact on the health of individuals (Sapolsky, 2005), their reproductive success (Clutton-Brock & Huchard, 2013), access to food resources (Kura, 2016) and safe resting places (Shackelford & Weekes-Shackelford, 2021). Moreover, the welfare of captive animals in a controlled environment might benefit from the understanding of the social dynamics; for instance, the effect of introducing a new member to a group or removing the alpha animal, which disturbs the stability of the current dominance hierarchy (Kura et al., 2015).

When studying social dominance relationships at the group level and the influence of intrinsic (such as age and physical characteristics (Kura et al., 2015)) or extrinsic factors (such as group composition and size, or past experiences of an animal (Dugatkin, 1997; Landau, 1951)) on their formation, researchers generally do not

directly observe dominance hierarchies as tangible entities. Instead, dominance hierarchies are latent structures that must be inferred from a network of dominance relationships among individuals. To achieve this, researchers typically collect and analyse a sequence of discrete occurrences (often sparse) of agonistic behaviours between individuals, which serves as the main source of information for revealing the underlying dominance hierarchy within that group. An observation minimally consists of the identities of the winner and the loser of an agonistic interaction, the time of occurrence and, optionally, the type of agonistic interaction.

Using an early definition by Schjelderup–Ebbe of a dominance relationship as a 'pattern of repeated, agonistic interactions between two individuals, characterized by a consistent outcome in favor of the same dyad member and a default yielding response of its opponent rather than escalation' (Schjelderup-Ebbe, 1922), such observations can be used to unravel the dominance relationships among the animals in a group. A plethora of sociometric measures have been proposed over the past decades that allow to quantify different properties of dominance relationships and potentially correlate them with intrinsic or extrinsic factors. These sociometric measures can be divided into two groups, depending on whether or not they use the temporal component of the observation. The first group relies on winner–loser (or actor–reactor) matrices of the agonistic interactions obtained by counting the number of (directed) agonistic interactions during one or several periods of observation. For this reason, they are often referred to as matrix-based methods (David, 1987; Hemelrijk et al., 2005). The second group, referred to as sequential methods (Albers & de Vries, 2001; Elo, 1978; Vilette et al., 2020), also incorporates the chronological order in which the interactions took place.

Matrix-based methods are further subdivided into three groups, according to the dimension of social dominance (dyadic, group or individual) they consider (Langbein & Puppe, 2004). At the dyadic level, dominance relationships can be characterized by simply counting the (relative) number of dyads in which there is a clear winner, that is where the number of wins of one animal significantly differs from the other (Lohse et al., 1982). As this level is closest to the definition of a dominance relationship, it is a natural point to start for an analysis. At the group level, indices such as the improved Landau *h* (De Vries, 1995) allow for expressing the degree of linearity of a dominance hierarchy. Finally, at the individual level, several sociometric measures have been proposed that use observed interactions to compute a dominance strength for each animal. This strength can be expressed by a cardinal score such as the Average Dominance Index (ADI) (Hemelrijk et al., 2005) or the normalized David score (David, 1987; De Vries et al., 2006), which can be reduced to dominance ranks. Alternatively, some methods can directly infer a dominance rank order from the interaction matrix, such as the Percolation and Conductance (P&C) method (Fujii et al., 2015). At the individual level, the methods mainly differ in how they account for the overall dominance level of an animal, whether they compensate for an outcome that occurred by chance, how dyads for which no interactions have been observed are treated and whether they rely on the counts of directed interactions per dyad or only consider the dominance direction (thus binary) per dyad.

By construction, matrix-based methods ignore the temporal component of dominance relationships. This makes them particularly ill suited to model the dynamic aspects of hierarchy formation. Even in studies that do not focus on the dynamics of social dominance, the presence of (potentially unknown) temporal effects can obscure the results. For example, when, due to the dynamics, the individual ranks of two animals flip in the middle of an extended observation period, the outcome of a pooled winner–loser matrix might be that the individuals cannot be ranked reliably. However, in reality, an unambiguous, yet time-dependent, ranking can exist. For example, Madge Pimentel et al. (2022) showed that dominance hierarchies in cichlid fish were observed to change dynamically in response to resource competition. Additionally, Sheng et al. (2024) found that dominance hierarchies in dairy cows flatten under heightened competition. One simple solution to overcome this problem is to partition the time domain into several intervals, construct winner-loser matrices per interval and compute the values for the sociometric measure multiple times. However, such a heuristic partitioning might not use the available data optimally, nor does it guarantee that, within each interval, the dominance relationships are stable.

Sequence-based methods provide an alternative to matrix-based methods by allowing to incorporate a temporal component. In addition to the outcome of an interaction, these methods also include the order in which the interactions took place. Most sequence-based methods described in literature and applied to model animal dominance hierarchies (to the best of our knowledge) are variants of the Elo rating (Elo, 1978), a scoring system originating from the world of chess which sequentially updates the scores of opponents based on the outcome of each game. Elo rating-based dominance analyses are gaining popularity through extensions such as the modified Elo rating (Newton-Fisher, 2017) and the steepness metric (Neumann & Fischer, 2023). The Elo rating considers dominance as a stochastic process in time (more precisely a Markov process (Hofman et al., 2020; Krifa et al., 2021)) that models the probabilities of the outcomes of agonistic interactions in time. It assigns a dominance score to each animal and models the winning probability of an animal against another as a function of the difference between the dominance scores of these animals. An initial dominance score is assigned to every animal at the start of the observation process. These initial values are often chosen to be equal (1000 is often used (Wooddell et al., 2017)) or can differ when, for instance, previous experiences correlated with dominance are available (Strauss & Holekamp, 2019a). When observing an agonistic interaction, the score of the winning animal is increased by an amount that is proportional to the winning probability of that animal against its opponent, with a proportionality constant *k* (called the update parameter) steering the magnitude of the updates. The dominance score of the loser is reduced by the same amount.

Variants of this basic method include the randomized Elo method (Sánchez-Tójar et al., 2018), where this process is repeated several times using random permutations (in time) of the observed

interactions and the obtained scores per permutation are averaged to obtain final dominance scores. The averaging is claimed to stabilize the Elo rating (Sánchez-Tójar et al., 2018), which is an advantage when the dominance relationships are stable. However, when temporal trends are present, this approach has the same disadvantages as the matrix-based sociometric measures. The Optimized Elo rating (Foerster et al., 2016) is an Elo variant in which the initial dominance scores, as well as the update parameter k, are considered estimatable parameters. Being a probabilistic model, these parameters can be estimated by likelihood maximization using an observed sequence of interactions. Interestingly, this approach allows for the use of the data to back-track the dominance scores to the start of the observation period and, using the update rules, compute the evolution of the scores as a function of time (Newton-Fisher, 2017). A further development of the Optimized Elo rating was proposed by Goffe et al. (2018). They use Bayesian inference techniques and assign a multivariate normal prior to the initial weights, however, at a considerable computational cost. This prior allows to stabilize the likelihood maximization process, especially for larger groups of animals with a limited number of observed interactions. As a side effect, this method also provides credibility intervals on the parameters, which are useful for uncertainty assessments as in the Glicko rating system (Glickman, 1995; Goffe et al., 2018).

Sequential methods allow to model time-dependent dominance scores and can be used to visualize the evolution of dominance over time. However, they are also criticized for several reasons. An important drawback is that they do not model dominance relationships at the dyadic level but at the individual level (Krahn et al., 2023), and thus do not align with the definition of Schjelderup–Ebbe. Moreover, in the context of a study on baboons, Gordon et al. (2022) claim the following issues concerning the original Elo rating method and its variants: (i) the rate of rank changes is slower compared to matrix-based ranking methods; (ii) the absence of observed interactions (absence of evidence) is treated as evidence that those interactions did not occur (evidence of absence). However, this assumption is often invalid in wild animal populations where often only a fraction of the interactions is observed. As a result, an animal can fall in rank by just having a few (observed) interactions, and changes in ranks can happen without any proof in the data for these changes; (iii) unexpected interactions result in disproportionate effects, as animals that have much-expected wins (e.g. a high-ranked animal wins against a low-ranked one) will have limited effects on their Elo scores as compared to a single unlikely loss (e.g. a high-ranked animal losing to a low-ranked one) against a much lower-ranked animal.

We argue that the Elo rating method conflates two separate processes. First, it attempts to model a stochastic process at the dyadic level. Second, it computes a sociometric measure that aims to quantify a property of the society at the individual level. By separating the stochastic model from the computation of the sociometric measure, a sequential model can be obtained that is able to address most of the criticisms given above. Therefore, in this paper, we propose a sequential matrix-based method called Animal Dominance Network (AniDomNet) that builds upon ideas from the extensions of the Elo rating to model the winning probabilities as a function of time. Figure 1 provides an overview of AniDomNet, showing its pipeline from interaction data to the inference of dominance hierarchies over time. As opposed to the existing Elo rating variants, it models the evolution of the pairwise winning probabilities as a function of time, rather than the evolution of the individual scores directly. The resulting model allows output of estimated pairwise winning probabilities at each point in time and can be plugged into many existing matrix-based sociometric measures. More precisely, optimizing multiple parameters in the update functions allows AniDomNet to determine the best rate of dyadic dominance changes. Additionally, the pairwise modelling of relationships with safeguards to prevent modelling relationships for animals that have never interacted tackles the problem of absence of evidence that Elo rating variants face when building dominance hierarchies.

Moreover, the use of the Cauchy distribution and the proposed updating scheme balance the disproportionate effects of unexpected events. Lastly, we propose an anomaly detector that is capable of pinpointing the most likely false positive events and identity switches. Together with the AniDomNet model, we propose a hierarchical ranking method that uses the estimated winning probabilities to rank the animals and also allows us to visually represent that ranking as a directed acyclic graph (DAG). In this paper, we mainly focus on groups with stable memberships (no inclusion or exclusion of individuals). However, a simple extension allows accommodating groups with changing membership. The proposed method is quantitatively compared to different existing Elo variants using a simulation study and a number of publicly available benchmark datasets from DomArchive (Strauss et al., 2022). An implementation of AniDomNet is publicly available at https://github.com/nusretipek/AniDomNet.

The remainder of this paper is organized as follows. Section 2 briefly introduces the sequential methods existing in the literature and the mathematical notations used. It then proceeds by introducing AniDomNet as the proposed stochastic method for the sequential dominance data and presents the experimental setup. Section 3 shows the quantitative benchmark results including AniDomNet as well as some qualitative results. A discussion of the advantages and shortcomings of the proposed method and possible applications is presented in this section. Lastly, the paper concludes with the most important insights and future research directions.

2 | MATERIALS AND METHODS

2.1 | Formal problem definition and Elo-based methods

The methods developed in this paper require edge-list format interaction data ordered in time, where each entry minimally consists of a winning animal and a losing animal. All interactions are assumed to be of the same type. Let $I = \{1, 2, \ldots, N\}$ be an index set labelling the individuals in a population of N animals. The dyad $(i, j) \in I \times I$, with $i \neq j$, represents an interaction between i and j with winner i

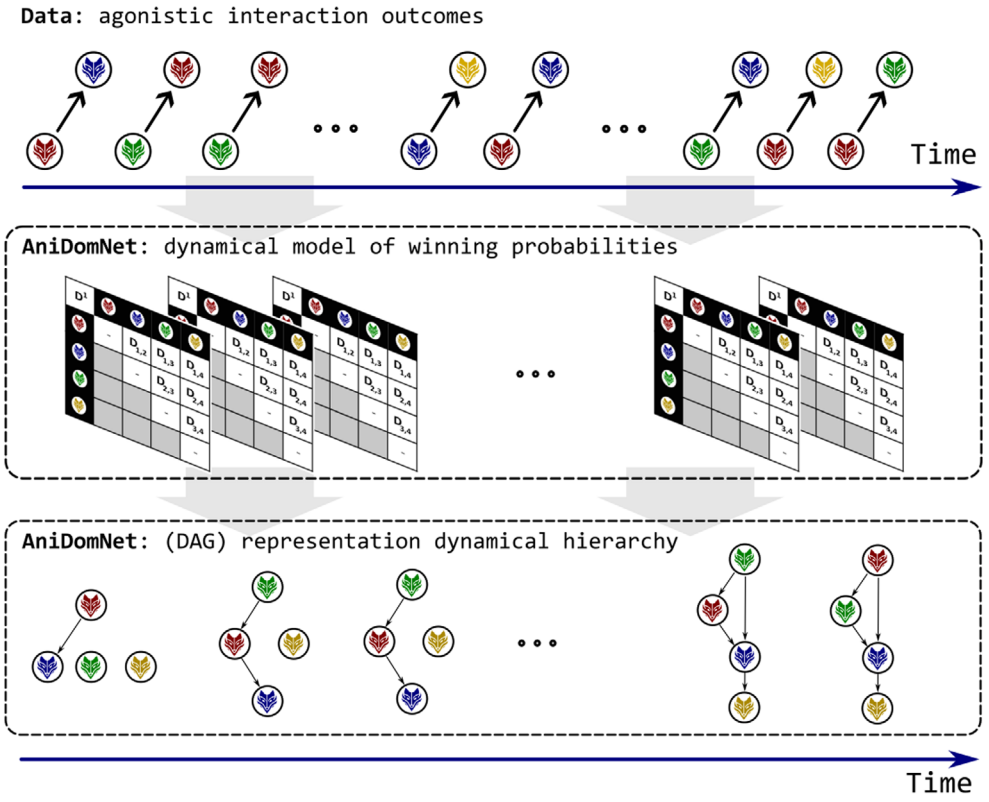

FIGURE 1 Overview of AniDomNet: The model uses agonistic interaction outcomes over time to infer dominance relationships. AniDomNet estimates the dynamical winning probabilities and represents hierarchical structures using a directed acyclic graph (DAG).

and loser j. For notational convenience, whenever i is used as an index, it will refer to a winning animal, whereas j will refer to a losing animal. Moreover, the index ℓ will be used to refer to an animal without the specific connotation of winning or losing. Additionally, we denote the *observed* interaction in the ordered edge-list at position (or time) t as (i_t, j_t). For example, $(i_t, j_t) = (2, 1)$ indicates that at time t animal 2 won in an agonistic interaction against animal 1. The sequence of interactions starting at $t = 1$ up to $t = T$ is denoted as:

$$X_T = \{(i_t, j_t) \mid t = 1, \ldots, T\}.$$

Let $P^t_{i,j}$ denote the probability that i defeats j at time t in a Bernoulli trial (with $P^t_{j,i} = 1 - P^t_{i,j}$). When considered as a function of t, the sequence of probabilities $P^t_{i,j}$ are the winning probabilities of a discrete-time stochastic process of Bernoulli trials. Both the different Elo variants as well as our proposed model will estimate these probabilities from the dataset X_T.

2.1.1 | The original Elo method

The original Elo method (Elo, 1978) assigns a score to each animal. Let $s^t_\ell \in \mathbb{R}$ represent this score for animal ℓ at time t. The winning probability at time t is calculated as:

$$P^t_{i,j} = \frac{1}{1 + e^{-\left(s^t_i - s^t_j\right)/\gamma}}, \quad (1)$$

where γ is a scale parameter, which is often set to 100. After observing (i, j) (thus an interaction where i defeats j), the scores s^t_i and s^t_j are updated as follows:

$$s^{t+1}_i = s^t_i + k\left(1 - P^t_{i,j}\right), \quad s^{t+1}_j = s^t_j - k\left(1 - P^t_{i,j}\right), \quad (2)$$

and the scores of all other animals are kept unchanged. The parameter $k \in \mathbb{R}^+$ is called the update parameter regulating the maximal gain or loss due to a single interaction and is often set experimentally (Neumann et al., 2011) (when not specified, we set $k = 100$ in this paper). The scores at time T are used as dominance measure to rank animals.

2.1.2 | The randomized Elo method

The randomized Elo method (Sánchez-Tójar et al., 2018) randomly permutes the observed ordered edge-list multiple times (n; $n = 1000$ is used in this paper) and the scores obtained at time T are subsequently averaged over the repetitions.

2.1.3 | Optimized Elo and Bayesian inference methods

The optimized Elo method (Foerster et al., 2016) considers the initial scores s^1_ℓ for $\ell = 1, \ldots, N$ and k as estimable parameters

and maximizes the likelihood of the data. The likelihood function $L(X_T; s_1^1, \ldots, s_N^1, k)$ is defined as the probability of observing the data X_T given the parameters $s_1^1, \ldots, s_N^1$ and k:

$$L(X_T; s_1^1, \ldots, s_N^1, k) = \prod_{t=1}^{T} P_{i_t j_t}^t. \quad (3)$$

Note that $P_{i_t j_t}^t$ represents the estimated probability that i_t will defeat j_t when they interact at time t and is thus computed after (recursively) applying the update formulae in Equation (2) to all dyads $(i_1, j_1), \ldots, (i_{t-1}, j_{t-1})$. For computational reasons, in practice the negative log-likelihood is minimized instead of maximizing L. The authors use the Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm (Fletcher, 1987) without constraints for solving this minimization problem. Additionally, Goffe et al. (2018) embed this model in a Bayesian framework and define a Gaussian prior on $s_1^1, \ldots, s_N^1$ and on k. Formally, they define the posterior of the parameters as:

$$p(s_1^1, \ldots, s_N^1, k, \sigma \mid X_T) \propto L(X_T; s_1^1, \ldots, s_N^1, k, \sigma)\, p(s_1^1, \ldots, s_N^1, k, \sigma), \quad (4)$$

where p represents a probability density function, $s_1^1, \ldots, s_N^1 \sim N(0, \sigma^2)$, $\sigma \sim N_+(0, 1)$ and $k \sim N(0, 1)$. Here, N_+ refers to the truncated Gaussian distribution. The mean of the posterior is used to estimate $s_1^1, \ldots, s_N^1$ and k.

2.2 | AniDomNet: Modelling winning probabilities

Just as the Elo method, AniDomNet models the winning probabilities of the animals involved in the interactions. However, instead of one score per animal, AniDomNet uses a real-valued $N \times N$ dominance matrix D^t, to characterize the dominance relationships at time t (see Figure 3 for an example). More precisely, we only use the strictly upper triangular part of D^t, thus elements $D_{i,j}^t$ for which $i < j$. Therefore, $D_{j,i}^t$ decreases if $i > j$ (see Equation (7)). Even though the elements of the lower triangular part are not used by the model, we set these elements equal to zero as this will simplify the notation in the following sections. The interpretation of D^t follows from the following definition of a winning probability $P_{i,j}^t$:

$$P_{i,j}^t = \begin{cases} \frac{1}{2} + \frac{\arctan\left(D_{i,j}^t\right)}{\pi} & , \text{if } i < j, \\ 1 - \left(\frac{1}{2} + \frac{\arctan\left(D_{i,j}^t\right)}{\pi}\right) & , \text{if } i > j. \end{cases} \quad (5)$$

Note that $\frac{1}{2} + \arctan\left(D_{i,j}^t\right)/\pi$ is the cumulative distribution function of the Cauchy distribution with location parameter 0 and scale parameter 1, evaluated at $D_{i,j}^t$. This means that $P_{i,j}^t = 0.5$ in case $D_{i,j}^t = 0$ and $P_{i,j}^t > 0.5$ only if $D_{i,j}^t > 0$. Therefore, i dominates j only if $D_{i,j}^t$ is positive and sufficiently large. On the other hand, j dominates i if $D_{i,j}^t < 0$. Equation (5) ensures that the computation is symmetric in the indices. The Cauchy distribution was selected for its heavy tails. As a result, unlikely but impactful events (such as a lower ranked animal winning against a highly ranked animal) are attributed a higher probability as compared to the Gaussian case, for instance. Nevertheless, it can be interchanged with any distribution that is symmetric about 0. When comparing Equations (1) and (5), it can be seen that $D_{i,j}^t$ is used as a substitute for $s_i^t - s_j^t$ and thus characterizes the dominance relationship between i and j in a similar, yet more flexible manner.

After observing (i, j), the entries of D^t are updated. To that end, we make a distinction between the primary dyad (i.e. the one in which both i_t and j_t are involved) and the auxiliary dyads (i.e. the ones in which only one of them is involved). The primary update is defined as:

(i) if $i < j$, then

$$D_{i,j}^{t+1} = D_{i,j}^t + \beta_1 \frac{W_{j,i}^t}{W_{i,j}^t}; \quad (6)$$

(ii) if $i > j$, then

$$D_{j,i}^{t+1} = D_{j,i}^t - \beta_1 \frac{W_{j,i}^t}{W_{i,j}^t}. \quad (7)$$

Here, $\beta_1 > 0$ is an update parameter similar to the parameter k in the Elo method. As a result of this update, the estimated winning probability of the winning animal will increase and the size of this increase is determined by β_1. Moreover, the effect of β_1 is modulated by the ratio of two elements from a time-dependent $N \times N$ weight matrix W^t. Just like D^t, the matrix W^t is updated using the following equations after observing (i, j):

$$W_{i,j}^{t+1} = W_{i,j}^t + \left(1 - P_{i,j}^t\right)\beta_3, \qquad W_{j,i}^{t+1} = \max\left(1, W_{j,i}^t - \left(1 - P_{i,j}^t\right)\beta_4\right). \quad (8)$$

To motivate this approach, we recall that Equation (2) uses $s_i^t + k\left(1 - P_{i,j}^t\right)$ to update the Elo scores. Here, $1 - P_{i,j}^t$ modulates the update parameter k by reducing the size of the update caused by interactions of which the outcome had a high probability. The intuition there is that, when i has a much higher score than j, an observed win of i against j should not affect the scores of i and j too much as the current scores already reflect this situation. Especially in the case of repeated attacks, this feature mitigates the updates. From Equation (8), it can be seen that $W_{i,j}^t$ accumulates winning probabilities, again with an update constant (β_3 for the winner and β_4 for the loser). Note that the truncation with the maximum operation prevents $W_{j,i}^{t+1}$ from becoming too small or even negative (even though in our experiments, this rarely happened even without such truncation). This modulation has three advantages over the Elo method: (i) it disconnects the direct link between $P_{i,j}^t$ and the modulation of the update parameter β_1, thus adding flexibility to the model as the effect of $P_{i,j}^t$ on the modulation can be estimated; (ii) winning and losing can be treated differently if the data would require so; and (iii) the gradually accumulated history of $P_{i,j}^t$ is used in the modulation as opposed to only its current value.

The primarily statistical arguments given to motivate the model structure are complemented by biological insights. The asymmetry in the effects of winning and losing (item (ii) above) is reviewed by Hsu et al. (2006) and has been demonstrated in several species (Camerlink et al., 2016). For instance, a series of papers showed that in sticklebacks the prior losing experience highly reduces the probability of winning while winning experience has only marginal effects (Bakker et al., 1989; Chase et al., 1994). Previous work also shows differences in the hormonal (e.g. testosterone) effects of winning and losing (Fuxjager et al., 2011). In several species, the effect of a winning experience disappears sooner than the effect of a losing experience (Hsu et al., 2006). Items (i) and (iii) can be linked to the work by Houston and McNamara (1999), which states that interaction outcomes can be described by the following cascade: experiences of previous interactions lead to (modified) expectations that subsequently affect the behaviour of the animals, and, as a consequence, also the outcome of an interaction. This two-stage cascade of expectations/experiences and behaviour also suggests a two-stage modelling approach. Using this analogy, W^t represents the expectations/experiences whereas D^t is an aggregate measure of behaviour.

The auxiliary updates will propagate the effect of the observed dyad (i,j) to interactions only involving one of i and j. First, we consider the relationships between the winner i and all other animals $\ell \notin \{i,j\}$. The corresponding updates are:

(i) if $\ell > i$, then

$$D^{t+1}_{i,\ell} = D^t_{i,\ell} + \beta_2 M^t_{i,\ell} \frac{W^t_{\ell,i}}{W^t_{i,\ell}}; \tag{9}$$

(ii) if $\ell < i$, then

$$D^{t+1}_{\ell,i} = D^t_{\ell,i} - \beta_2 M^t_{i,\ell} \frac{W^t_{\ell,i}}{W^t_{i,\ell}}. \tag{10}$$

Here, $\beta_2 > 0$ is an auxiliary update parameter. This update propagates the winner's bonus to the relationships with the other animals, by increasing the probability that i will also win from the animals not involved in (i,j). Moreover, this update formula uses a binary interaction matrix M^t in which $M^t_{i,j} = M^t_{j,i} = 1$ only if i and j have interacted before time t and $M^t_{i,j} = 0$ otherwise. This approach avoids auxiliary updates affecting the winning probabilities between animals that never interacted before. At time $t = 1$, the matrix M^1 is a zero matrix.

Secondly, consider the relationships between the loser j and all other animals $\ell \notin \{i,j\}$:

(i) if $\ell < j$, then

$$D^{t+1}_{\ell,j} = D^t_{\ell,j} + \beta_2 M^t_{j,\ell} \frac{W^t_{j,\ell}}{W^t_{\ell,j}}; \tag{11}$$

(ii) if $\ell > j$, then

$$D^{t+1}_{j,\ell} = D^t_{j,\ell} - \beta_2 M^t_{j,\ell} \frac{W^t_{j,\ell}}{W^t_{\ell,j}}. \tag{12}$$

This auxiliary update propagates the loser's tax to the relationships with the other animals, by increasing the probability that j will also lose from the animals not involved in (i,j).

2.3 | AniDomNet: Parameter estimation

AniDomNet depends on initial values for the matrices D ($N(N-1)/2$ values as only the strictly upper diagonal part is needed) and W ($N^2 - N$ values) and the parameters β_1, β_2, β_3 and β_4. Initial values and the parameters can be estimated by maximizing the following likelihood function:

$$L(X_T; D^1, W^1, \beta_1, \beta_2, \beta_3, \beta_4) = \prod_{t=1}^{T} P^t_{i_t j_t} \tag{13}$$

where $P^t_{i_t j_t}$ is computed using Equation (5).

The parameter space can be constrained by several inequalities. A first inequality we consider is $\beta_1 \geq \theta_1 \beta_2$ to enforce that primary updates have a larger impact on D^t than auxiliary (indirect) updates. We set $\theta_1 = 3$ in our experiments. Moreover, the constraint $\beta_3 \geq \theta_2 \beta_4$ can be added to enforce asymmetry in the updating of W^t. We experimentally found little effect on the fit of the model, but adding this constraint (with $\theta_2 = 3$) was beneficial for the runtime of the (numerical) optimization process (see experimental result for a comparison of the results without this constraint). Lastly, we set $W^1_{i,j} = 1$ for every i and j and minimize the negative logarithm of Equation (13) resulting in the following optimization problem:

$$\begin{aligned} \underset{D^1,\beta_1,\beta_2,\beta_3,\beta_4}{\text{minimize}} \quad & \sum_{t=1}^{T} -\log\left(P^t_{i_t j_t}\right) \\ \text{subject to} \quad & \beta_1 \geq \theta_1 \beta_2 \\ & \beta_3 \geq \theta_2 \beta_4 \\ & \beta_1, \beta_2, \beta_3, \beta_4 \geq 0 \end{aligned} \tag{14}$$

To solve this minimization problem, the sequential least squares programming optimization method (SLSQP) is utilized as it allows to minimize non-convex loss functions with constraints (Kraft, 1988; Virtanen et al., 2020). As the problem is non-convex, the solution might depend on the choice of the initial values of the parameters. Therefore, we perform the minimization 100 times, sampling each $D^1_{i,j}$ from a uniform distribution over $[-1.0,\ 1.0]$. Initial values for β_1, β_2, β_3 and β_4 were resampled uniformly from $[0,\ 0.1]$ until values were obtained satisfying the inequality constraints. After each optimization run, the negative log-likelihood loss is stored. In the end, the random initial parameters that yield the lowest negative log-likelihood are retained. To increase the computational efficiency, we further: (i) Omit elements $D_{i,j}$ for couples (i,j) for which no interactions are observed during the experiment. As the binary interaction matrix $M^t_{i,j} = 0$ for all values t, these values are not relevant to both the model and the optimization process. In the resulting model, $D^t_{i,j} = 0$ for all such couples. (ii) The objective function in Equation (14) and the update rules in Equations (6–12) are implemented in the

low-level language C. The code is coupled with the SLSQP implementation of SciPy (v.1.11.3) (Virtanen et al., 2020) to obtain a user-friendly end product that can be called from Python. The code and the data are open source in the GitHub repository referred to in the introduction.

2.4 | AniDomNet light

In Section 2.2, the ratio $\frac{W^t_{ji}}{W^t_{ij}}$ in Equations (6–12) was motivated as an alternative to the Elo rating update which uses $1 - P^t_{i,j}$ for this purpose. A simplified version of AniDomNet can thus be obtained by changing the update rules in Equations (6–12) into an Elo method like update. For the primary updates, the equations become

$$D^{t+1}_{i,j} = D^t_{i,j} + \beta_1\left(1 - P^t_{i,j}\right), \quad D^{t+1}_{j,i} = D^t_{j,i} - \beta_1\left(1 - P^t_{j,i}\right). \tag{15}$$

Similarly, for the auxiliary updates, in Equations (9) and (10) the ratio $\frac{W^t_{\ell,i}}{W^t_{i,\ell}}$ is replaced with $1 - P^t_{\ell,i}$, whereas in Equations (11) and (12), the ratio $\frac{W^t_{j,\ell}}{W^t_{\ell,j}}$ is replaced with $1 - P^t_{j,\ell}$. These changes effectively allow to eliminate the matrix W^t from the method leading to a significant simplification. In the results, we describe the impact of this simplification on the performance of AniDomNet and further elaborate on this in Supporting Information.

2.5 | Quantitative evaluation of AniDomNet

Sequential methods, such as AniDomNet, enable the accurate prediction of future outcomes based on past interactions over time. In our experiments, the parameters for AniDomNet were estimated once using the complete dataset. However, after the initial estimation, the model dynamically updates based on the previous interactions, X^t, to predict the probability of the next interaction ($P^{t+1}_{i_t j_t}$). This evolving probability, although not directly reflected in the previously established hierarchy D^t, plays a critical role in calculating both the accuracy and loss of the model. This approach is consistent with similar methodologies used in optimized Elo and Bayesian inference models (Foerster et al., 2016; Goffe et al., 2018). During the optimization process, the model's parameters are expected to converge by incorporating all available data, meaning that the accuracy measured here is not akin to traditional machine learning accuracy, where distinct training and test sets are used. Instead, the goal is to measure how well the model can predict the interaction outcomes and thus how well it represents the underlying (dominance) phenomena. Consequently, AniDomNet continuously refines its predictions and optimizes its parameters iteratively across the entire dataset.

Given a sequence of interactions, the accuracy of a model can be computed as the ratio of the number of interactions for which the outcome was predicted correctly, to the total number of observed interactions. Formally, given a sequence of interactions $X_T = \{(i_t, j_t)\}_{t=1}^T$, the optimal values for the parameters of the optimized Elo method and AniDomNet can be estimated by maximizing the likelihood (Equations 3 and 13) or using Bayesian inference.

The resulting parameters can subsequently be used to predict the outcome of each interaction in X_T.

Note that $P^t_{i_t,j_t}$ is the predicted probability that, at time t, animal i_t dominates j_t. To transfer this probability into a crisp forecast, we use the common convention that $P^t_{i_t j_t} \geq 0.5$ means that the model predicts that i_t dominates j_t at time t. Following our notational conventions, $(i_t, j_t) \in X_T$ means that it was observed that i_t won against j_t at time t (and is thus the ground truth). The agreement between the crisp forecasts and the observed outcomes at $t = 1, \ldots, T$ is referred to as the accuracy and is computed as follows:

$$\text{Acc} = \frac{1}{T}\sum_{t=1}^{T} 1\left[P^t_{i_t j_t} \geq 0.5\right], \tag{16}$$

where $1_{\left[P^t_{i_t j_t} \geq 0.5\right]} = 1$ if $P^t_{i_t j_t} \geq 0.5$ and 0 otherwise. Note that the models involved are dynamic, which means that P^t will evolve over time by applying the update rules. Similarly, the negative log-likelihood (generally referred to as the loss) can be computed for each model as:

$$\mathscr{L} = \sum_{t=1}^{T} -\log\left(P^t_{i_t j_t}\right). \tag{17}$$

When interpreting the accuracy and the negative log-likelihood, it should be stressed that both performance measures only indicate how well the model can describe the observed (agonistic) interactions. These measures do not distinguish between outcomes that are the result of elucidations of the dominance hierarchy and outcomes that are the result of social dynamics not related to the dominance hierarchy (for instance, aggression because of leverage (Lewis, 2002)). Moreover, more subtle interactions that are indicative of the dominance hierarchy are often overlooked in datasets focusing on high-intensity agonistic interactions. Claims regarding accurate modelling of the dominance hierarchy should be made with care.

2.5.1 | Evaluation on real-world data

To compare the predictive capabilities of our model with the Elo variants, both performance measures were computed on all sequential (benchmarking) datasets collected in DomArchive (Strauss et al., 2022). Table 1 provides some details on the datasets used for comparing the models.

AniDomNet can also be used as an anomaly detector. After fitting the model, for each (i_t, j_t) in a dataset, the probability $P^t_{i_t j_t}$ can be computed. When $P^t_{i_t j_t} < \epsilon$, the observed interaction can be flagged as an anomaly. Especially for (fully) automated, sensor-based technologies for the detection of agonistic interactions, detection errors are to be expected (Hrolenok et al., 2018; Ipek et al., 2023), though anomalies may also arise from various causes. These anomalies should be the focus of further investigation, rather than simply being treated as erroneous readings. For instance, the identity of

TABLE 1 Descriptive dataset information; dataset name, species, sex, captivity, number of animals and number of interactions, * not included in DomArchive, but available in Vilette et al. (2020).

Dataset name	Species	Sex	Captive	No. of animals	No. of interactions
Foerster2016a (Foerster et al., 2016)	*Pan troglodytes* (Chimpanzee)	M	No	22	2741
Foerster2016b (Foerster et al., 2016)	*Pan troglodytes* (Chimpanzee)	F	No	44	1015
Franz2015a (Franz et al., 2015)	*Papio cynocephalus* (Yellow baboons)	M	No	61	4118
Franz2015b (Franz et al., 2015)	*Papio cynocephalus* (Yellow baboons)	M	No	28	1667
Franz2015c (Franz et al., 2015)	*Papio cynocephalus* (Yellow baboons)	M	No	36	2387
Franz2015d (Franz et al., 2015)	*Papio cynocephalus* (Yellow baboons)	M	No	53	4464
Franz2015e (Franz et al., 2015)	*Papio cynocephalus* (Yellow baboons)	M	No	52	3281
Langley2018 (Langley et al., 2018)	*Phasianus colchicus* (Common pheasant)	M	Yes	18	367
McCune2019a (McCune et al., 2019)	*Aphelocoma wollweberi* (Mexican jays)	MF	No	7	201
McCune2019c (McCune et al., 2019)	*Aphelocoma wollweberi* (Mexican jays)	MF	No	8	163
McCune2019d (McCune et al., 2019)	*Aphelocoma wollweberi* (Mexican jays)	MF	No	8	151
McCune2019e (McCune et al., 2019)	*Aphelocoma wollweberi* (Mexican jays)	MF	No	7	80
McCune2019f (McCune et al., 2019)	*Aphelocoma wollweberi* (Mexican jays)	MF	No	8	53
Strauss2019a (Strauss & Holekamp, 2019b)	*Crocuta crocuta* (Spotted hyena)	F	No	23	1065
Strauss2019b (Strauss & Holekamp, 2019b)	*Crocuta crocuta* (Spotted hyena)	F	No	35	1848
Strauss2019c (Strauss & Holekamp, 2019b)	*Crocuta crocuta* (Spotted hyena)	F	No	29	1913
Strauss2019d (Strauss & Holekamp, 2019b)	*Crocuta crocuta* (Spotted hyena)	F	No	151	9096
Vilette2020a* (Vilette et al., 2020)	*Chlorocebus pygerythrus* (Vervet monkey)	MF	No	61	8301
Vilette2020b (Vilette et al., 2020)	*Chlorocebus pygerythrus* (Vervet monkey)	MF	No	41	2980

Note: The McCune2019f dataset contains only 53 interactions for 8 animals, which is below the minimum number of interactions suggested for the calculation of a dominance hierarchy (Sánchez-Tójar et al., 2018). Thus, the results should be interpreted with caution.

the animals involved in an interaction can be incorrect or an interaction is wrongfully identified as being agonistic. Also, human observations are subject to subjectivity and errors while interpreting interactions (Tuyttens et al., 2014). In those cases, $P^t_{i_t j_t}$ is expected to be small, which can help to identify potential anomalies for further review. To investigate the application of AniDomNet as an anomaly detector, a subset of the data described in Ipek et al. (2023) is used. In this work, the authors developed a computer vision pipeline that allows to detect agonistic interactions (and the identity of the animal involved) among group-housed lactating female rabbits. In a sequence of 1760 hours of video, 1633 interactions were detected and used to fit AniDomNet. Out of these interactions those for which $P^t_{i_t j_t} < 0.5$ were inspected by a human expert.

2.5.2 | Evaluation on simulated data

While the results obtained on real-world datasets can be used to numerically compare AniDomNet with the Elo variants, the true dominance hierarchy is unknown for these datasets. To evaluate how well AniDomNet can recover the true hierarchy, a simulation experiment is set up. Four scenarios are considered. In each scenario, the underlying hierarchy is represented by a directed graph (see Figure 2 for an example). From these graphs, 800 interactions are sampled by randomly selecting two nodes that are connected by a directed path. The outcome of the interaction respects the dominance relation imposed by the graph. With a small probability, the outcome is flipped to mimic observer or sensor errors or momentary animal incentives (Ipek et al., 2023; Tuyttens et al., 2014) that oppose the dominance hierarchy. This probability depends on the pair of animals and is for every pair sampled from a uniform distribution over $[0,\ \alpha]$. This means that the average probability of flipping over all pairs approaches $\alpha/2$. The scenarios differ, however, in the following ways: Scenarios S1 and S2 simulate a stable hierarchy, using $\alpha = 0.2$ and $\alpha = 0.3$, respectively, that is represented by the transitive, directed graph *State 1* in Figure 2, excluding the connection in the red edge. Scenario S3 represents a non-transitive setting, where the hierarchy is given be *State 1*, but now including the red edge.

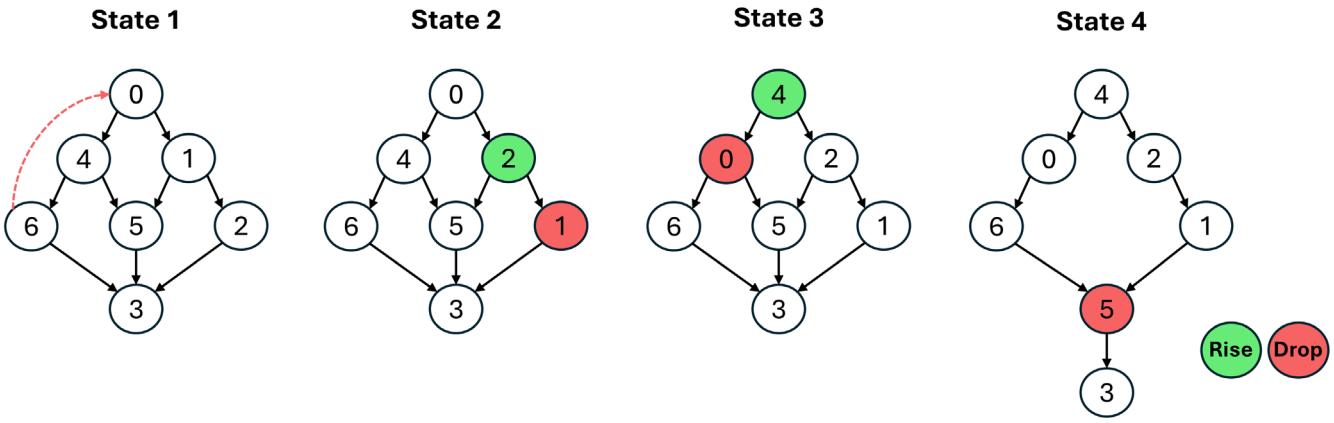

FIGURE 2 Illustration of the *true dominance hierarchy* and its changes over 800 interactions for the simulated datasets. The hierarchy updates every 200 interactions, with rank changes shown by colour-coded nodes: Green for rises and red for drops. Three major shifts occur: A mid-hierarchy change ($t = 200$), an alpha replacement ($t = 400$) and a rank drop without replacement ($t = 600$). A cycle is marked by a red dashed arrow (State 1), from Animal 6 to Animal 0. The transitivity assumption holds across all hierarchies, except the cycle.

Scenario S4 represents an evolving case where the initial hierarchy is represented by State 1, but changes to State 2 after every 200 observed interactions, and to States 3 and 4 after another 200 observed interactions. Each scenario is repeated 100 times.

2.6 | Post-processing AniDomNet: Ranking and representation as a DAG

AniDomNet is designed to model the evolution of dominance relationships over time. After fitting the model, the matrix D^t contains the information that allows the dominance relationships at time t to be described. Note that also at small values of t, for instance, D^1 at the start of the experiment, AniDomNet allows the state of the dominance relationships to be estimated as it backpropagates the interactions at later time points back to the start of the experiment. However, D^t (which is a strictly upper triangular matrix) can be challenging to use directly. Interestingly, several quantities can be derived from D^t that are much simpler to interpret. In the next paragraphs, we describe how to use D^t to compute the dominance score and dominance rank for each animal at time t. Additionally, we describe how to visually represent the dominance hierarchy as a directed acyclic graph (DAG).

From the definition of the model, it follows that when $D^t_{i,j} \gg 0$, animal i dominates j. Moreover, when $D^t_{i,j} \ll 0$, the opposite is true and when $D^t_{i,j} \approx 0$, the relationship is unsettled. Also, the magnitude of $D^t_{i,j}$ can be used as a measure of the strength of the dominance of one animal over another. Therefore, by summing the entries of D^t that involve an animal ℓ, a score can be obtained that characterizes the overall dominance of ℓ. However, due to the construction of D^t, one needs to deal carefully with the signs of the elements that are added. When ℓ is the row index, then $D^t_{\ell,j} \gg 0$ means that ℓ dominates j and $D^t_{i,\ell} \ll 0$ means that ℓ is dominated by i. Therefore, a dominance score can be obtained by adding all elements of the ℓ-th row and subtracting all elements from the ℓ-th column to obtain a score s_ℓ for the ℓ-th animal:

$$s^t_\ell = \sum_{j=1}^{N} D^t_{\ell,j} - \sum_{i=1}^{N} D^t_{i,\ell}, \tag{18}$$

where we complete the strictly upper triangular matrix D^t into a regular matrix by setting $D^t_{i,j} = 0$ for all $i \geq j$. The resulting scores $s^t_1, \ldots, s^t_N$ can be interpreted as dominance scores and can be used to rank the animals.

The obtained scores can be complemented by a representation of the dominance relationships as a directed graph $G = (V, E)$, where each vertex in $V = \{1, \ldots, N\}$ represents an animal and the directed edges $E \subset V \times V$ indicate the (dyadic) dominance relationships. The (strictly upper triangular part of) matrix D^t can be used to define E, where $(i,j) \in E$ if $D^t_{i,j} > \epsilon$ and $(j,i) \in E$ if $D^t_{i,j} < -\epsilon$, where ϵ is a threshold that distinguishes between settled and unsettled relationships. Note that, following our definition of D^t, $D^t_{i,j} = 0$ for all couples (i,j) that never interacted. Therefore, the absence of an edge either means that: (i) there is an absence of evidence of a relationship between i and j; or (ii) the relationship is unsettled.

Despite its simplicity, the graph construction procedure above often leads to visualizations that are overcrowded and hard to interpret (Douglas et al., 2017). A more structured, simplified representation can be obtained by exploiting some properties of dominance relations. Mathematically, a dominance relation is a binary relation on $N \times N$. Strict total order relations are a specific type of binary relations that are irreflexive, antisymmetric, transitive and complete. In a dominance context, irreflexivity means no animal dominates itself. Antisymmetry means that whenever i dominates ℓ, then ℓ does not dominate i. Transitivity means that whenever i dominates ℓ and ℓ dominates j, then i also dominates j. Completeness means that there are no unsettled relationships. A dominance relation that satisfies these properties allows us to linearly rank the animals, meaning that the individuals constitute a total order. When dropping the completeness property, a strict partial order relation is obtained, which can be represented as a directed acyclic graph (DAG). It should be noted that dropping the

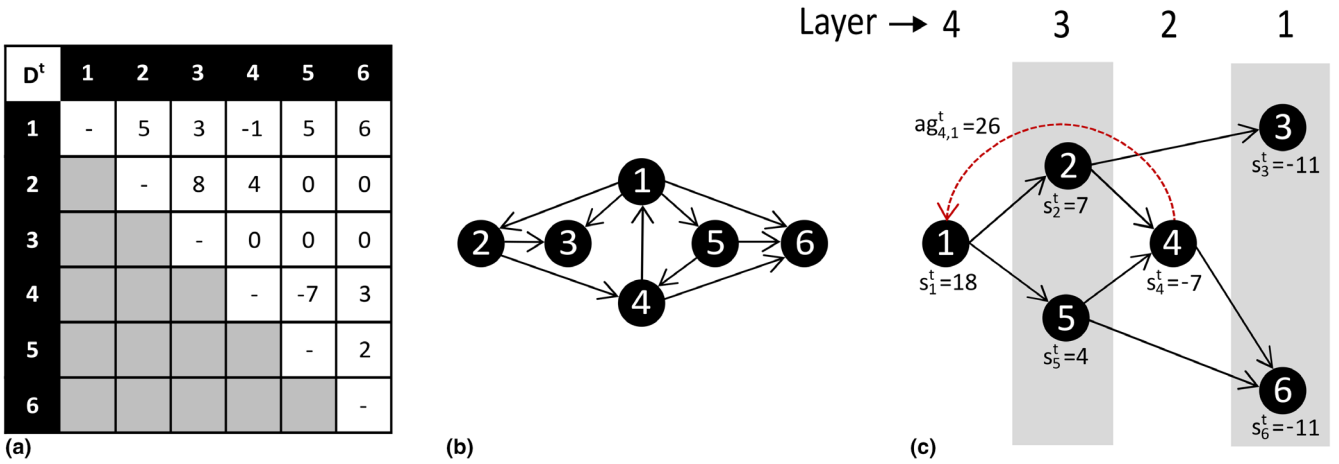

FIGURE 3 (a) Matrix D^t represents a hypothetical example of pairwise dominance relationships among six animals with Cauchy-distributed variables. (b) Directed graph of dominance relationships based on matrix D^t assuming $\epsilon = 0$. Each directed edge represents a dominance relationship from dominating animal to the subordinate one. In this hypothetical example, two cycles are present: $1 \to 2 \to 4 \to 1$ and $1 \to 5 \to 4 \to 1$. (c) Directed Acyclic Graph (DAG) representation of the dominance hierarchy. The cycle from $4 \to 1$ is broken based on the social agony and layers are formed according to bottom-up layout.

completeness property implies that animals can be incomparable, which can be interpreted as the absence of a dominance relationship between the animals involved. On the other hand, when the completeness of the underlying dominance hierarchy can be assumed but, due to limitations of the observation process or the statistical analysis, the relation cannot be observed, the lack of evidence can also be represented using a DAG. However, in this case, the DAG is only a tool used to express uncertainty about a relationship rather than the absence of one.

Our DAG construction method mostly adheres to the second point of view, where the absence of a relationship represents uncertainty about its direction. Additionally, our representation assumes that the underlying dominance relationship is transitive and that violations of transitivity are due to limitations in the observation process or the statistical modelling. The starting point is the graph G introduced earlier. In the first step to obtain a DAG, cycles are broken by removing a limited number of edges from G. In the first iteration, all fixed-length cycles of length $k = 3$ are detected. The task of finding all cycles of length less than or equal to k can be completed in $O\big((c+n)(k-1)d^k\big)$ steps, where n is the number of vertices, c is the total number of cycles discovered, and d is the average vertex degree, using a depth-first search algorithm with backtracking, as proposed by Gupta and Suzumura (2021). For edges that are present in at least one cycle, the social agony is computed, a measure originating from social network analysis (Sun et al., 2017). For an edge (i,j), the social agony is defined as:

$$\mathrm{ag}_{i,j}^t = \max\left(0, s_j^t - s_i^t + 1\right). \tag{19}$$

When the social agony is large, it means that even though (i,j) is an edge and thus according to G animal i dominates j, the score s_j^t of j is much larger than that of i. The edge (i,j) thus contradicts the order that would be obtained when using the scores to rank the animals. Therefore, the edge with the largest agony is removed from G (see Figure 3). Subsequently, the procedure is repeated using the updated graph until all cycles of length $k = 3$ are removed. In the next iteration, all cycles of length $k = 4$ are considered. This procedure is iterated until $k = N$. The resulting graph, denoted G^*, is thus guaranteed to be free of cycles. Lastly, we provide a layered layout procedure for G^*. Similar to ADAGIO (Douglas et al., 2017), the layout is constructed bottom-up. The first layer consists of all animals that do not dominate any other animal according to G^*, that is those nodes that do not have any outgoing edge. The second layer consists of all animals that only dominate animals from the first layer, that is all nodes in G^* that only have outgoing edges to nodes in the first layer. This procedure is repeated for the remaining layers. In general, the m-th layer consists of nodes with at least one outgoing edge to a node in layer $m-1$ and all outgoing remaining edges only to layers $m-2, \ldots, 1$. The algorithms are presented formally in the Supporting Information.

2.7 | Qualitative analyses with AniDomNet

AniDomNet can be used to gain insight into the evolution of the dominance relationships as a function of time. Interestingly, at any point in time t, the matrix D^t contains all information on the dominance relationships extracted from the data. Moreover, D^t encodes information at the pairwise level, which resembles the format needed by the so-called matrix-based methods, especially those that focus on binary dominance relationships rather than the number of occurrences of a specific dyad. Recall that i dominates j if either $D_{i,j}^t > \epsilon$ or $D_{j,i}^t < -\epsilon$. Using these inequations, a binary dominance matrix can be constructed that is used by dominance measures such as the improved Landau h (De Vries, 1995) or Kendall's K (Appleby, 1983) (linearity indices). In an additional experiment, we compute the linearity of the dominance hierarchy using the improved Landau h index. We

compared the indices computed from the interaction matrix created directly from X^t with those from D^t, where the relationships were binarized based on $D^t_{i,j} > \epsilon$ or $D^t_{j,i} < -\epsilon$ using $\epsilon = 0$. These computations were performed for each dataset in DomArchive and graphically presented in the Supporting Information.

Additionally, the dominance hierarchies were visualized using the DAG layout procedure described in Section 2.6, primarily focusing on the Langley2018 (Langley et al., 2018) dataset. To stress the impact of the visualization procedure on the conclusions that tentatively can be drawn, we compare our DAG layout method with visualizations derived from optimized Elo, randomized Elo and Bayesian inference methods.

3 | RESULTS AND DISCUSSION

3.1 | Quantitative analysis

3.1.1 | Evaluation on real-world data

Table 2 presents the (weighted) accuracy (Equation 16) and loss (Equation 17) of the sequential methods and the corresponding runtime (see Supporting Information for detailed hardware/software information). In terms of predictive power, the original Elo method performs the worst, whereas Bayesian inference is the best among the Elo variants; however, at the cost of a significant increase in runtime. Bayesian inference results in an average loss of 226 and weighted accuracy of 97.04% over all datasets. AniDomNet outperforms all Elo variants in terms of both average loss and weighted accuracy with 125 and 98.20%, respectively.

The significant overall reduction of average loss and increase in accuracy suggests that AniDomNet has a better fit to the observed interactions. This improvement indicates that dominance relationships derived from the model will also be more reliable and accurate.

Sequential methods in the literature often use a burn-in period to overcome the lack of a reliable estimate of the initial state of the dominance relationships. Note that in Table 2, no burn-in period is used, as estimating the initial state of the dominance relationships is part of the optimized Elo, Bayesian inference (Foerster et al., 2016) methods and AniDomNet model. For this reason, the results are not directly comparable with to those that use a burn-in period (Foerster et al., 2016; Goffe et al., 2018). Moreover, Vilette et al. (2020) used a train and test set approach to evaluate sequential methods. Although this approach implies that train and test sets show similar dynamics (which was not verified), we use the train (Vilette_2020a) and test (Vilette_2020b) datasets from Vilette et al. (2020), along with their approach, to evaluate AniDomNet for completeness. Hence, the state of Elo ratings in Bayesian inference and state of D^T are preserved without additional adjustments during the test phase. The Bayesian inference method results in an accuracy of 83.69% (we were not able to reproduce the 82.2% in Table 2 (Vilette et al., 2020)), while AniDomNet reaches 85.8% for the test dataset.

For validation of anomaly detection, 1633 automatically annotated agonistic interactions from Ipek et al. (2023) are used. AniDomNet results in 359 anomalies where $P^t_{i_t,j_t} < 0.5$. Among these detections, 290 are either identity switches of the two animals involved or false positives. In this dataset, AniDomNet is capable of pinpointing mistakes with a precision of 80.8%. Albeit simple, the anomaly detection technique based on the contribution of an observation to the loss seems to be able to detect (at least a significant amount of) anomalies. Removing these anomalies can improve the validity of conclusions drawn about the dominance relationships.

AniDomNet Light was benchmarked on the same datasets. This ablation study reveals that the simplified version resulted in accuracy of 96.01%, which is moderately better than the optimized Elo method, but substantially worse than the full AniDomNet. For a more elaborate comparison, see Supporting Information. We conclude that the matrix W^t and its updating rules contribute substantially to the overall performance.

3.1.2 | Evaluation on simulated data

Table 3 shows the loss and accuracy obtained for the four scenarios included in the simulation study. Comparing AniDomNet with Optimized Elo and Bayesian Inference, a notable reduction in the loss can be observed. This is not unexpected as AniDomNet is designed to be more flexible (it has more parameters than the Elo variants). For scenarios S1 and S2, the difference in accuracy is very small (the marginal accuracy gain likely stems from AniDomNet's extended parametrization). All approaches obtain an accuracy that approaches the average noise level of $\alpha / 2$. This is expected as State 1 represents a (transitive) dominance hierarchy where the winning probabilities can be modelled using the Elo variants. By selecting the Elo scores such that:

$$s_0 > s_1 = s_4 > s_2 = s_5 = s_6 > s_3,$$

with a sufficiently large margin for the inequalities, an optimal accuracy can be obtained. For scenario S3, it can be seen that the introduction of a cycle has a negligible impact on AniDomNet, whereas it results in a strong drop of the accuracy for the Elo variants. This scenario shows that the increased flexibility of AniDomNet allows it to outperform the Elo variants when modelling non-transitive dominance hierarchies. Lastly, in dynamic scenario S4 (where hierarchies are transitive but change in time), a slight drop in accuracy is observed for the Elo variants compared to S1, whereas AniDomNet is less affected. Lastly, it can be observed that the setting of θ_2 has limited effect on the results of AniDomNet. Overall, the loss and accuracy are slightly better when $\theta_2 = 0$, but computation time almost doubles when $\theta_2 = 0$, compared to when $\theta_2 = 3$.

Figure 4a shows the DAGs of scenarios S1 and S2 at $t = 800$, obtained by AniDomNet's bottom-up graph construction algorithm. Figure 4b shows the result of the graph construction procedure for scenario S3 at $t = 800$ with and without cycle removal. In each case, the recovered graph is equal to the one used to simulate the data (we

TABLE 2 Benchmark results: A comprehensive comparison of four sequential methods: Original Elo, Optimized Elo, Bayesian Inference and AniDomNet (the proposed model), across various datasets from DomArchive (Strauss et al., 2022).

Dataset		Original Elo			Optimized Elo			Bayesian inference			AniDomNet		
Name	Interaction numbers	Time (s)	Loss	Accuracy	Time (s)	Loss	Accuracy	Time (s)	Loss	Accuracy	Time (s)	Loss	Accuracy
Foerster_2016a	2741	<1	401	94.78%	26	372	95.37%	750	172	98.61%	548	116	98.72%
Foerster_2016b	1015	<1	394	83.35%	78	293	88.77%	432	256	91.03%	602	114	96.16%
Franz_2015a	4118	<1	951	90.51%	166	799	92.35%	2823	503	96.43%	3477	376	97.09%
Franz_2015b	1667	<1	345	92.86%	70	263	95.20%	570	196	96.46%	394	136	97.24%
Franz_2015c	2387	<1	520	92.33%	48	452	92.92%	1006	223	97.82%	704	176	97.61%
Franz_2015d	4464	<1	727	94.22%	242	612	95.30%	3233	320	98.10%	2047	314	98.14%
Franz_2015e	3281	<1	723	91.92%	203	645	92.62%	2066	356	97.04%	1508	235	97.65%
Langley_2018	367	<1	99	89.65%	12	65	92.92%	115	47	96.46%	206	12	98.91%
McCune_2019a	201	<1	49	92.04%	2	30	96.52%	55	4	100.00%	127	2	99.50%
McCune_2019c	163	<1	47	87.12%	1	34	92.03%	48	6	99.39%	48	2	99.39%
McCune_2019d	151	<1	47	90.07%	1	34	94.70%	46	7	98.68%	305	9	98.01%
McCune_2019e	80	<1	25	91.25%	1	14	93.75%	36	3	100.00%	94	3	98.75%
McCune_2019f	53	<1	16	86.79%	1	5	98.11%	33	4	100.00%	91	1	98.11%
Strauss_2019a	1065	<1	216	92.68%	44	139	95.96%	347	63	98.69%	1451	30	99.06%
Strauss_2019b	1848	<1	411	92.26%	280	330	93.89%	825	98	99.35%	2635	35	99.51%
Strauss_2019c	1913	<1	286	95.35%	178	166	98.43%	791	76	99.37%	2836	11	99.79%
Strauss_2019d	9096	<1	2232	91.29%	1756	1735	94.07%	26,560	897	98.01%	31,195	431	98.86%
Vilette_2020b	2979	<1	1090	84.77%	272	917	88.12%	1185	834	89.69%	3452	256	97.25%
*Average**	2088	<1	477	91.52%	188	384	93.63%	2273	226	97.04%	2873	125	98.20%

Note: For each method, the table includes the time in seconds required to complete the calculations, the negative log-likelihood loss ($\mathscr{L}$), and the accuracy as a proxy for predictive power for the outcome of X^{t+1}. The term *Average** represents the arithmetic average of the number of interactions, time and loss, while the weighted average of accuracy is calculated with weights proportional to the number of interactions in the dataset. Overall, AniDomNet demonstrates the lowest loss and the highest weighted accuracy compared to the other methods. Despite this superior performance in terms of loss and accuracy, AniDomNet requires only slightly more computation time on average.

TABLE 3 Simulation results: AniDomNet, Optimized Elo and Bayesian Inference ratings across four simulation scenarios.

Scenario			AniDomNet ($\theta_2 = 0$)			AniDomNet ($\theta_2 = 3$)			Optimized Elo		Bayesian inference	
Name	T	α	Loss	Accuracy	Time (s)	Loss	Accuracy	Time (s)	Loss	Accuracy	Loss	Accuracy
S1: State 1	800	0.2	204	91.16%	45	214	90.62%	21	278	90.05%	267	90.06%
S2: State 1	800	0.3	269	87.16%	45	279	86.06%	22	349	85.07%	335	85.26%
S3: State 1 + cycle	800	0.2	216	90.38%	46	221	89.89%	20	334	83.68%	321	84.79%
S4: Dynamic	800	0.2	207	91.01%	71	219	90.43%	29	307	88.44%	264	89.47%

Note: The parameter α indicates the average noise level during interaction sampling. Two settings for AniDomNet are considered, with $\theta = 0$ and $\theta = 3$, which effectively remove the constraint on the W matrix update parameters. The loss and the accuracy metrics are average of 100 randomly generated simulation datasets. The time refers to the average run time for fitting AniDomNet per simulation.

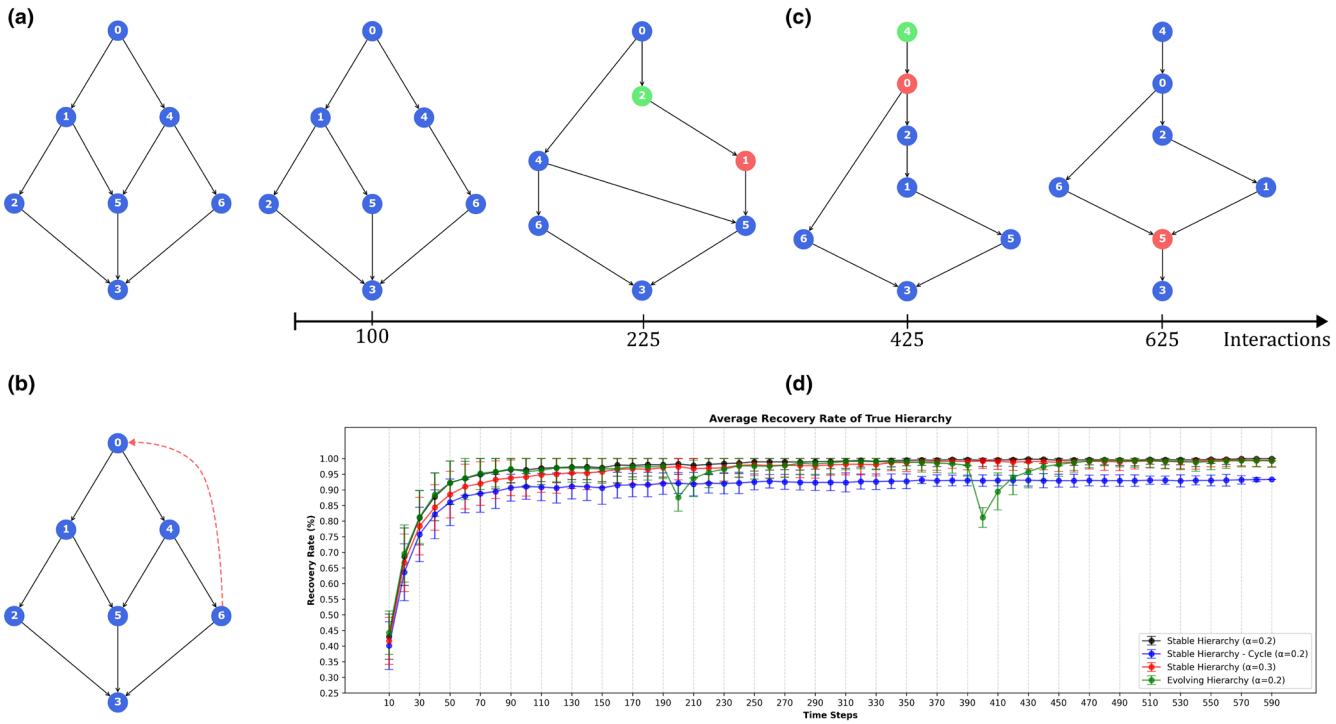

FIGURE 4 (a) Bottom-up DAG illustration based on pairwise relationships of *S1: State 1* at $t = 800$, same for $\alpha = 0.2$ and $\alpha = 0.3$. (b) Bottom-up DAG illustration based on pairwise relationships of *S1: State 1 + Cycle* before removal of the cycle. (c) Bottom-up DAG illustration based on pairwise relationships of *Evolving Hierarchy* in continuous time scale. The hierarchy before $t = 100$ and after $t = 625$ can be extrapolated to $t = 0$ and $t = 800$, respectively. (d) Recovery rate of the true dominance hierarchy in simulated settings using AniDomNet. Lines indicate the average recovery rate over 100 simulations. The number of interactions required for convergence depends on α—a higher α increases the interactions needed to infer the true hierarchy. In the evolving hierarchy setting, changes in dominance structure highlight AniDomNet's adaptability (green line).

show the results of a representative simulation). Figure 4c shows constructed DAGs at multiple time points for scenario S4. Note that the hierarchy switches every 200 interactions. This example shows that the model and DAG construction procedure allows to adapt to a changing hierarchy. In this simulation, 25 interactions after the changes happened, the switches already resulted in updated hierarchies that resemble the changes. It should be noted that this is the result of one simulation. To study this effect more generally, Figure 4d shows the hierarchy recovery rate (the Hamming distance on the adjacency graphs of the true hierarchy and the one obtained with AniDomNet) averaged over all simulations. It can be seen here that, for the evolving hierarchy, the average recovery rate exceeds 90% over about 100 observed interactions. When a change in the dominance hierarchy is induced, the drop in recovery of the average recovery rate disappears after observing about 50 interactions. Care should be taken, however, not to overgeneralize these findings to other cases as the number of observations/time needed heavily depends on the number of animals involved, the type of change and so on.

3.2 | Qualitative analysis

Figure 5 shows the evolution of the improved Landau *h* (a linearity index) computed on the observed interaction matrix (interactions are accumulated over time) and computed using the DAG from D^t after estimating the parameters of AniDomNet once using the whole dataset. Both converge to 0.615 at the end of the observation period. On the other hand, the improved Landau *h* calculated using only the interactions up to *t* shows a rise in the linearity over time compared to the relatively stable linearity based on the matrix D^t. These, at first sight conflicting observations, can be attributed to the way in which *h* handles unknown relationships. When the number of observations is limited (small *t*), most relationships are considered unsettled due to a lack of observation data. However, as time passes, the interaction matrix fills up and Landau *h* increases. This phenomenon is considered undesirable as the apparent rise of linearity is merely an artefact of the lack of data. Moreover, disentangling a change in *h* due to lack of data on one hand and a change due to a true change in the linearity of the dominance hierarchy becomes very challenging. On the other hand, by design, AniDomNet is not affected by this lack of data in the early stages of an experiment. A change in *h* computed using D^t is therefore likely to be more indicative of the stability and linearity of the hierarchy.

The convergence also suggests that D^t effectively captures the global dominance structure that the matrix-based methods aim to model at the end of the observation period. Figure 5a–c displays the directed graph obtained from D^t for $t = 100$, 200 and 300, respectively. The red edges indicate the cycles broken based on the social agony.

Figure 6 shows several DAG representations of the estimated dominance hierarchy at the end of the experiment (time *T*). Figure 6a uses the Elo scores from the optimized Elo method to linearly order the animals and represents the result as a graph. Similarly, the randomized Elo method (Figure 6b) uses the mean Elo scores to rank the animals. Here, the random permutation of the data gives rise to an empirical score distribution per animal. For animals 6 and 8, the 90% confidence intervals derived from these distributions overlap. Therefore, they were assigned the same rank in the visualization. The Bayesian inference method (Figure 6c) uses the posteriors on the Elo scores. The resulting DAG is constructed by ordering the animals according to the mean of the posterior. Moreover, the credibility intervals based on the 10th and 90th percentiles are derived from these posteriors (using the original code implementation) (Goffe et al., 2018). When the intervals of two animals overlap, the animals are considered to be incomparable and are not connected by an edge. Lastly, Figure 6d displays a bottom-up DAG constructed from D^T using the procedure described in Section 2.6. Comparing these representations, it is clear that the method used has a significant impact on how the dominance relationships are represented. Clearly, the DAG derived from the optimized Elo ratings (as in Figure 6a) is

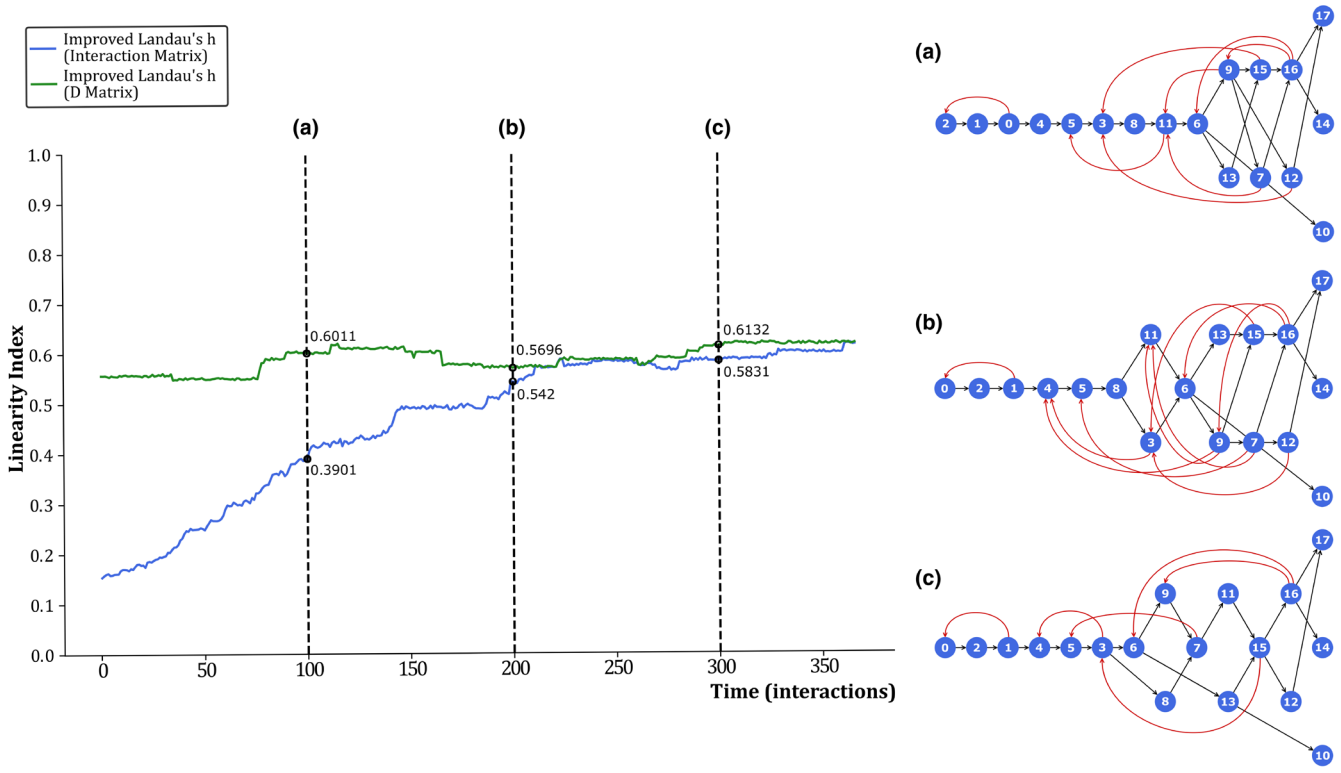

FIGURE 5 Evolution of linearity for the Langley2018 (Langley et al., 2018) dataset using the improved Landau *h* based on observed interaction data and the improved Landau *h* based on matrix D^t. At interaction points 100 (a), 200 (b) and 300 (c), dominance hierarchies with bottom-up ranking are given. The black directed arrows show the final DAG, whereas the red directed arrows indicate the cycles that were broken during the construction process.

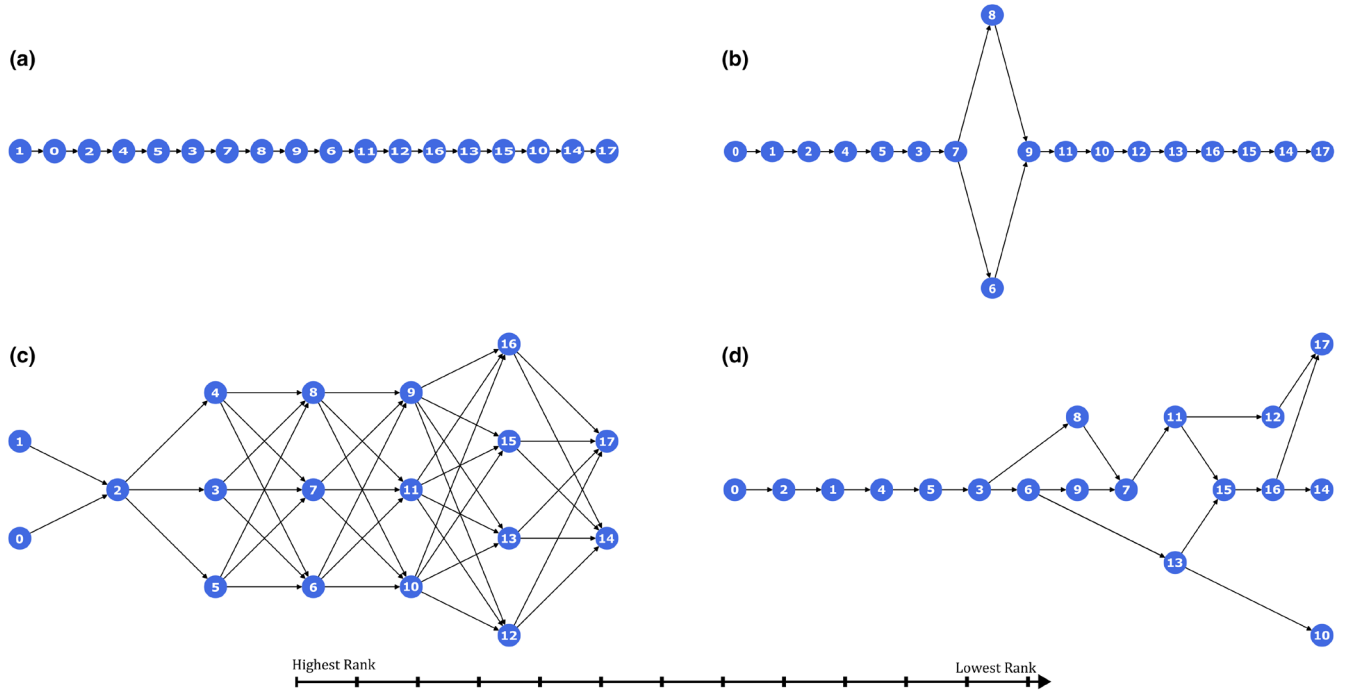

FIGURE 6 Visualization of DAG representations of the Langley2018 dataset (Langley et al., 2018) using sequential methods. Dominance hierarchies are calculated with all observed interactions, hence representing the state at the end of the study (*T*). For illustrative purposes, some animals are assigned the same rank based on overlapping dominance measures (mean Elo or likelihood). An animal is placed at lower rank level if its dominance measure is lower and does not share overlapping confidence/credibility intervals with all animals at the current level. For animals within the same rank (see b–d), individual dominance measures decrease from bottom to top. For instance, the mean Elo score of animal 6 is higher than that of animal 8 in (b). (a) Optimized Elo method: Results in a linear order. (b) Randomized Elo method: Results in confidence intervals based on 1000 original Elo method repetitions ($\alpha = 0.1$). (c) Bayesian inference method: Partial pooling allows for the estimation of uncertainty around the mean Elo estimate of individual animals (10th and 90th percentiles, namely Q1 and Q9, are used). (d) AniDomNet: Bottom-up DAG based on pairwise relationships.

linear. We argue that this approach can be misleading because it suggests that the true dominance relationship is linear. However, there is not necessarily robust evidence in the data that supports a strongly linear dominance hierarchy. The approaches used to obtain Figure 6b,c do try to capture some degree of uncertainty in the scores and translate the uncertainty in the ordering of two animals into incomparability. Even though the resulting graphs do relax the overall assumption of linearity, they are based on the Elo method that does assume a linear ordering, which again is too strict for some species or animal groups. Linearity constraints that are inherent to these methods also lead to spurious orderings among animals. For instance, animals 10, 14 and 17 did not win any agonistic interaction (in Langley2018 (Langley et al., 2018)); their final positions in the hierarchy are only influenced by defeats. For instance, even if animal 10 had fewer defeats compared to all other animals within this group, ranking it higher in the hierarchy despite not having a single win would be counter-intuitive and highlights the issue of interpreting the absence of evidence as evidence of dominance (as is the case in Figure 6b,c). The DAG in Figure 6d, which is based on AniDomNet, relies less heavily on the assumption of linearity as the probabilistic model underlying AniDomNet does not assume linearity. Nevertheless, the algorithmic choices made to translate D^t into a graph do assume that the underlying dominance hierarchy is at least representable as a DAG. Indeed, the cycle-breaking procedure was motivated by interpreting some cycles as being the result of modelling errors and are therefore removed. Moreover, the incomparability of animals is rather the result of an absence of evidence of the nature of the relationship as opposed to a proof of them being equal. Nevertheless, given enough interactions, the DAG construction procedure of AniDomNet would converge to a linear hierarchy as Elo rating and its variants. In case this property is unwanted, the cycle removal procedure should be omitted from the DAG construction.

Social dynamics are complex due to factors such as the environment, coalitions and other latent factors (Chase et al., 2002). Coupled with a limited number of observed interactions, these complexities result in some pairwise relationships lacking sufficient evidence to determine dominance. The original Elo and optimized Elo methods almost always boil down to a perfectly linear DAG unless more than one animal in the dataset neither wins nor loses. In such cases, a statistical deficiency arises in the optimized Elo method, where the marginal likelihoods are flat.

Although the randomized Elo method and the Bayesian inference method might relax the perfectly linear DAG by quantifying uncertainty around the estimated ratings, the assumption of transitivity combined with completeness results in a DAG representation without evidence. Compared to the other sequential methods,

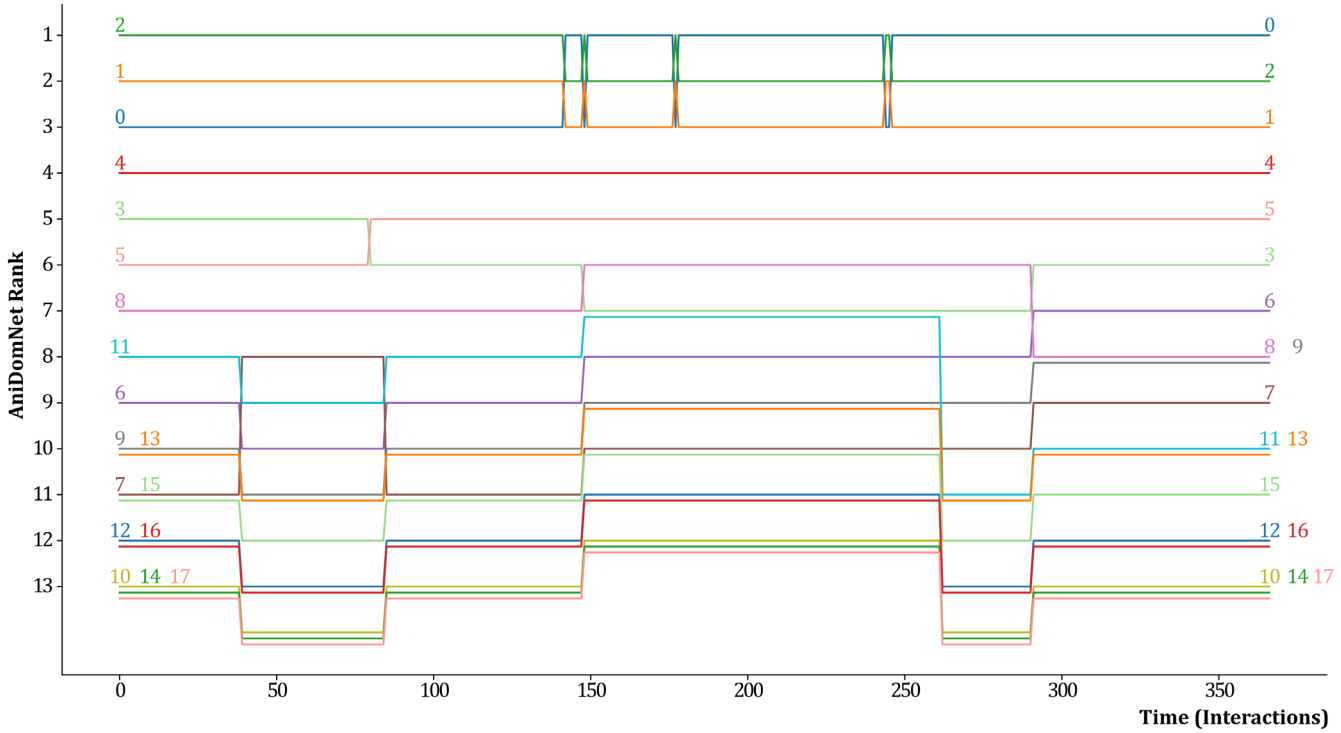

FIGURE 7 Illustration of the AniDomNet hierarchy rank evolution over time for all animals in the Langley2018 (Langley et al., 2018) dataset. The coloured numbers corresponding to the lines indicate the animal identification. The AniDomNet layer rank represents the ordinal rank, ranging from 1 to 13, where 1 is the highest level in the hierarchy. The ranks are updated after each observed interaction, serving as a proxy for time.

AniDomNet does not assume completeness; hence, the pairwise unsettled dominance relationships are not represented without evidence. In fact, the DAG constructed on the basis of the D^T matrix provides a comprehensive view by allowing for the representation of both linear and non-linear hierarchies. Moreover, the social agony-based cycle breaking improves the consistency in the representation of dominance hierarchies while preserving transitivity. Consequently, the AniDomNet DAG representation is expected to be closer to the underlying dominance hierarchy.

Figure 7 shows the evolution of the layer rank for each animal, as determined from AniDomNet's DAG. At certain time points, there are additional levels of hierarchy emerging whereas the overall ranks appear to be stable over time. Between $t = 0$ and $t = 150$, the layer rank of the top-3 animals (0, 1 and 2) is stable. After $t = 150$, there is a change in the hierarchy where animal 0 rises to the first rank with only short-interval interruptions during the settlement. Furthermore, animal 7 ranked higher early on ($t = 40$ to $t = 75$) due to wins against higher-ranking animals. In comparison, the original Elo rating using observed data until $t = 75$ rank animal 7 higher than animal 0, even though both animals have an equal number of wins and losses. Overall, a graphical illustration of ranks over time is useful to represent the social dynamics and rank stability. The AniDomNet layer ranks strive for a balance between flexibility when there is evidence of change in hierarchy and stability when there is an absence of evidence.

3.3 | Limitations and extensions

The work presented in this paper generalizes Elo rating by modelling dominance at the pairwise level. The previous sections have shown that modelling at the pairwise level increases the flexibility, for example in non-transitive dominance hierarchies. It comes, however, at the cost of an increased complexity of the model. Also, whereas the scores resulting from the Elo rating can be directly interpreted, the output of the AniDomNet model is harder to interpret and should be post-processed. The DAG construction procedure proposed to post-process the modelling outcome will eliminate edges that result in cycles to obtain a representation of the hierarchy as a directed acyclic graph. Therefore, this post-processing method should be used with caution if transitivity cannot be assumed.

An important aspect that has not been considered in this study relates to the occurrence of demographic events (births, deaths and migration) or, even more generally, *sudden events* that have a strong impact on the dominance hierarchy on a short time scale (Strauss & Holekamp, 2019a; Strauss & Shizuka, 2022). Similar to Elo rating, the updating mechanism of AniDomNet is a global mechanism that assumes that dominance hierarchies (and the resulting winning probabilities) change gradually in time. Societies where such gradual changes are mixed with sudden events that abruptly shake up the hierarchy remain hard to model. A study on the ability to model

such a mix is beyond the scope of this paper, but considering the structure of both Elo rating and the AniDomNet model, the ability of these models to adapt to sudden changes requires large values for the update parameters. Such values will, however, lead to an overresponse to gradual changes or temporal outbursts that do not reflect the dominance hierarchy. This may in turn lead to overestimation of the dynamics. In the case of demographic events, a careful implementation would allow to extend or reduce the matrices D and W as a function of time (see Supporting Information for a more elaborate discussion and implementation), but this extension would not account for the potentially impactful and abrupt change of the hierarchy.

The experimental evaluation showed that AniDomNet can improve upon Elo rating. When the true dominance hierarchy is known and non-transitive, it is clear that Elo rating is not flexible enough and that the added flexibility provided by the AniDomNet model is of added value. Also, the graph construction procedure allows for a more nuanced view of the dominance hierarchy. However, the simulation experiment included in this paper is limited to a small group of animals. We did not systematically compare AniDomNet with Elo rating in large groups nor groups where the rules that decide which animals will interact are more complex (such as simulation models involving spatial components). Such analysis could allow a user to make more informed decisions on when to use AniDomNet and to assess, for instance, how many interactions should be observed to make reliable claims on the dominance hierarchy. To decide on that, we can only advise a user to set up a simulation experiment that reflects the situation (in terms of group size and an assumed structure of the hierarchy) to create the recovery plots such as those in Figure 4d. A simple example of how such a study can be set up is added in the Supporting Information.

Lastly, it should be noted that AniDomNet has no mechanisms to explicitly model ternary relationships, where, for instance, bystander effects or the formation of coalitions are taken into account.

4 | CONCLUSIONS

As an alternative to sequential methods that estimate dominance relationships from pairwise interaction data, this study presented the AniDomNet model. Particularly, AniDomNet addresses several shortcomings of the current sequential methods, resulting in a better fit to the underlying observed interaction data, reaching an overall weighted accuracy of 98.20%. Consequently, the dominance hierarchies derived from the pairwise relationships are closer to the true dominance hierarchies of the animal group under study. Existing Elo-based methods use the scores assigned to each animal both as a sociometric measure and as latent variables in the probabilistic model describing the evolution of the dominance relationships over time, which constrains the types of relations that can be modelled accurately. AniDomNet decouples the modelling and the derived sociometric measure, leading to a more flexible modelling approach.

AUTHOR CONTRIBUTIONS

Nusret Ipek: Conceptualization; methodology; software; writing—original draft; visualization; investigation; formal analysis; data curation. Frank A. M. Tuyttens: Writing—review and editing; supervision. Bernard De Baets: Conceptualization; writing—review and editing; supervision. Jan Verwaeren: Conceptualization; methodology; writing—review and editing; validation; supervision.

ACKNOWLEDGEMENTS

This study was funded by the Ghent University Special Research Fund (BOF), by the Flemish Government under the 'Onderzoeksprogramma Artificiële Intelligentie (AI) Vlaanderen'. Part of the data (group-housed rabbits) was collected as part of the project with reference DO/SID/DWZ/OO/18/04 funded by the Government of Flanders, Department Environment, Section Animal Welfare. The computational resources (Stevin Supercomputer Infrastructure) and services used in this work were provided by the VSC (Flemish Supercomputer Center), funded by Ghent University, FWO and the Flemish Government—department EWI. We thank Alessandro Barbini for his invaluable assistance with software testing.

CONFLICT OF INTEREST STATEMENT

The authors declare no competing interests.

PEER REVIEW

The peer review history for this article is available at https://www.webofscience.com/api/gateway/wos/peer-review/10.1111/2041-210X.70118.

DATA AVAILABILITY STATEMENT

All source code and data during the current study are available on Zenodo: https://zenodo.org/records/16026360 (Ipek et al., 2025) as well as in the GitHub repository: https://github.com/nusretipek/AniDomNet.

ORCID

Nusret Ipek https://orcid.org/0000-0002-4507-0200
Frank A. M. Tuyttens https://orcid.org/0000-0002-1348-218X
Bernard De Baets https://orcid.org/0000-0002-3876-620X
Jan Verwaeren https://orcid.org/0000-0001-9978-501X

REFERENCES

Albers, P. C., & de Vries, H. (2001). Elo-rating as a tool in the sequential estimation of dominance strengths. *Animal Behaviour*, *61*(2), 489–495.
Appleby, M. C. (1983). The probability of linearity in hierarchies. *Animal Behaviour*, *31*(2), 600–608.
Bakker, T. C., Bruijn, E. F.-D., & Sevenster, P. (1989). Asymmetrical effects of prior winning and losing on dominance in sticklebacks (*Gasterosteus aculeatus*). *Ethology*, *82*(3), 224–229.
Camerlink, I., Peijnenburg, M., Wemelsfelder, F., & Turner, S. P. (2016). Emotions after victory or defeat assessed through qualitative behavioural assessment, skin lesions and blood parameters in pigs. *Applied Animal Behaviour Science*, *183*, 28–34.

Chase, I. D., Bartolomeo, C., & Dugatkin, L. A. (1994). Aggressive interactions and inter-contest interval: How long do winners keep winning? *Animal Behaviour*, *48*(2), 393–400.

Chase, I. D., Tovey, C., Spangler-Martin, D., & Manfredonia, M. (2002). Individual differences versus social dynamics in the formation of animal dominance hierarchies. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(8), 5744–5749.

Clutton-Brock, T. H., & Huchard, E. (2013). Social competition and selection in males and females. *Philosophical Transactions of the Royal Society, B: Biological Sciences*, *368*(1631), 20130074.

David, H. A. (1987). Ranking from unbalanced paired-comparison data. *Biometrika*, *74*(2), 432–436.

De Vries, H. (1995). An improved test of linearity in dominance hierarchies containing unknown or tied relationships. *Animal Behaviour*, *50*(5), 1375–1389.

De Vries, H., Stevens, J. M., & Vervaecke, H. (2006). Measuring and testing the steepness of dominance hierarchies. *Animal Behaviour*, *71*(3), 585–592.

Douglas, P. H., Ngomo, A.-C. N., & Hohmann, G. (2017). A novel approach for dominance assessment in gregarious species: Adagio. *Animal Behaviour*, *123*, 21–32.

Dugatkin, L. A. (1997). Winner and loser effects and the structure of dominance hierarchies. *Behavioral Ecology*, *8*(6), 583–587.

Elo, A. E. (1978). *The rating of chessplayers, past and present*. Arco Pub.

Fletcher, R. (1987). *Practical methods of optimization*. Wiley. Number 2. c. in A Wiley-Interscience publication.

Foerster, S., Franz, M., Murray, C. M., Gilby, I. C., Feldblum, J. T., Walker, K. K., & Pusey, A. E. (2016). Chimpanzee females queue but males compete for social status. *Scientific Reports*, *6*(1), 35404.

Franz, M., McLean, E., Tung, J., Altmann, J., & Alberts, S. C. (2015). Self-organizing dominance hierarchies in a wild primate population. *Proceedings of the Royal Society B: Biological Sciences*, *282*(1814), 20151512.

Fujii, K., Jin, J., Shev, A., Beisner, B., McCowan, B., & Fushing, H. (2015). *Perc: Using percolation and conductance to find information flow certainty in a direct network*. R package version 0.1.3.

Fuxjager, M. J., Montgomery, J. L., & Marler, C. A. (2011). Species differences in the winner effect disappear in response to post-victory testosterone manipulations. *Proceedings of the Royal Society B: Biological Sciences*, *278*(1724), 3497–3503.

Glickman, M. E. (1995). The glicko system. *Boston University*, *16*(8), 9.

Goffe, A. S., Fischer, J., & Sennhenn-Reulen, H. (2018). Bayesian inference and simulation approaches improve the assessment of elo-ratings in the analysis of social behaviour. *Methods in Ecology and Evolution*, *9*(10), 2131–2144.

Gordon, J. B., Jansen, D., Learn, N. H., & Alberts, S. C. (2022). *Ordinal dominance rank assignments: Protocol for the amboseli baboon research project*. Amboseli baboon research project.

Gupta, A., & Suzumura, T. (2021). Finding all bounded-length simple cycles in a directed graph. *arXiv preprint*, arXiv:2105.10094. https://doi.org/10.48550/arXiv.2105.10094

Hemelrijk, C., Wantia, J., & Gygax, L. (2005). The construction of dominance order: Comparing performance of five methods using an individual-based model. *Behaviour, 142*, 1037–1058.

Hofman, A. D., Brinkhuis, M. J., Bolsinova, M., Klaiber, J., Maris, G., & van der Maas, H. L. (2020). Tracking with (un)certainty. *Journal of Intelligence*, *8*(1), 10.

Houston, A. I., & McNamara, J. M. (1999). *Models of adaptive behaviour: An approach based on state*. Cambridge University Press.

Hrolenok, B., Balch, T., Byrd, D., Roberts, R., Kim, C., Rehg, J. M., Gilliland, S., & Wallen, K. (2018). Use of position tracking to infer social structure in rhesus macaques. In *Proceedings of the fifth international conference on animal-computer interaction* (pp. 1–5). Association for Computing Machinery.

Hsu, Y., Earley, R. L., & Wolf, L. L. (2006). Modulation of aggressive behaviour by fighting experience: Mechanisms and contest outcomes. *Biological Reviews*, *81*(1), 33–74.

Ipek, N., Tuyttens, F., De Baets, B., & Verwaeren, J. (2025). AniDomNet: A sequential pairwise model for inferring dynamic animal dominance hierarchies (Version v0.1.0) [Computer software]. Zenodo. https://doi.org/10.5281/ZENODO.16026360

Ipek, N., Van Damme, L. G., Tuyttens, F. A., & Verwaeren, J. (2023). Quantifying agonistic interactions between group-housed animals to derive social hierarchies using computer vision: A case study with commercially group-housed rabbits. *Scientific Reports*, *13*(1), 14138.

Kraft, D. (1988). *A software package for sequential quadratic programming*. Forschungsbericht-Deutsche Forschungs und Versuchsanstalt fur Luft und Raumfahrt.

Krahn, J., Foris, B., Weary, D. M., & von Keyserlingk, M. A. (2023). Invited review: Social dominance in dairy cattle: A critical review with guidelines for future research. *Journal of Dairy Science*, *106*(3), 1489–1501.

Krifa, A., Spinelli, F., & Junca, S. (2021). On the convergence of the Elo rating system for a Bernoulli model and round-robin tournaments. PhD thesis, Université Côte D'Azur.

Kura, K. (2016). Mathematical modelling of dominance hierarchies. PhD thesis, City, University of London.

Kura, K., Broom, M., & Kandler, A. (2015). Modelling dominance hierarchies under winner and loser effects. *Bulletin of Mathematical Biology*, *77*, 927–952.

Landau, H. G. (1951). On dominance relations and the structure of animal societies: I. Effect of inherent characteristics. *The Bulletin of Mathematical Biophysics*, *13*, 1–19.

Langbein, J., & Puppe, B. (2004). Analysing dominance relationships by sociometric methods—A plea for a more standardised and precise approach in farm animals. *Applied Animal Behaviour Science*, *87*(3–4), 293–315.

Langley, E. J., van Horik, J. O., Whiteside, M. A., & Madden, J. R. (2018). Group social rank is associated with performance on a spatial learning task. *Royal Society Open Science*, *5*(2), 171475.

Lewis, R. J. (2002). Beyond dominance: The importance of leverage. *The Quarterly Review of Biology*, *77*(2), 149–164.

Lohse, H., Ludwig, R., & Röhr, M. (1982). *Statistische Verfahren für Psychologen, Pädagogen und Soziologen*. Volk und Wissen Volkseigener Verlag.

Madge Pimentel, I., Vitt, S., & Thünken, T. (2022). Impact of genetic relatedness and food competition on female dominance hierarchies in a cichlid fish. *Animal Behaviour*, *184*, 55–70.

McCune, K. B., Jablonski, P., Lee, S.-i., & Ha, R. R. (2019). Captive jays exhibit reduced problem-solving performance compared to wild conspecifics. *Royal Society Open Science*, *6*(1), 181311.

Neumann, C., Duboscq, J., Dubuc, C., Ginting, A., Irwan, A. M., Agil, M., Widdig, A., & Engelhardt, A. (2011). Assessing dominance hierarchies: Validation and advantages of progressive evaluation with elo-rating. *Animal Behaviour*, *82*(4), 911–921.

Neumann, C., & Fischer, J. (2023). Extending bayesian elo-rating to quantify the steepness of dominance hierarchies. *Methods in Ecology and Evolution*, *14*(2), 669–682.

Newton-Fisher, N. E. (2017). Modeling social dominance: Elo-ratings, prior history, and the intensity of aggression. *International Journal of Primatology*, *38*(3), 427–447.

Sánchez-Tójar, A., Schroeder, J., & Farine, D. R. (2018). A practical guide for inferring reliable dominance hierarchies and estimating their uncertainty. *Journal of Animal Ecology*, *87*(3), 594–608.

Sapolsky, R. M. (2005). The influence of social hierarchy on primate health. *Science*, *308*(5722), 648–652.

Schjelderup-Ebbe, T. (1922). Contributions to the social psychology of the domestic chicken. *Zeitschrift für Psychologie*, *88*, 225–252.

Shackelford, T. K., & Weekes-Shackelford, V. A. (2021). *Encyclopedia of evolutionary psychological science*. Springer.

Sheng, K., Foris, B., Krahn, J., Weary, D. M., & von Keyserlingk, M. A. (2024). Redefining dominance calculation: Increased competition flattens the dominance hierarchy in dairy cows. *Journal of Dairy Science*, *107*(9), 7286–7298.

Strauss, E. D., DeCasien, A. R., Galindo, G., Hobson, E. A., Shizuka, D., & Curley, J. P. (2022). Domarchive: A century of published dominance data. *Philosophical Transactions of the Royal Society B*, *377*(1845), 20200436.

Strauss, E. D., & Holekamp, K. E. (2019a). Inferring longitudinal hierarchies: Framework and methods for studying the dynamics of dominance. *Journal of Animal Ecology*, *88*(4), 521–536.

Strauss, E. D., & Holekamp, K. E. (2019b). Social alliances improve rank and fitness in convention-based societies. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(18), 8919–8924.

Strauss, E. D., & Shizuka, D. (2022). The dynamics of dominance: Open questions, challenges and solutions. *Philosophical Transactions of the Royal Society B*, *377*(1845), 20200445.

Sun, J., Ajwani, D., Nicholson, P. K., Sala, A., & Parthasarathy, S. (2017). Breaking cycles in noisy hierarchies. In *Proceedings of the 2017 ACM on web science conference* (pp. 151–160). ACM Digital Library.

Tuyttens, F., de Graaf, S., Heerkens, J. L., Jacobs, L., Nalon, E., Ott, S., Stadig, L., Van Laer, E., & Ampe, B. (2014). Observer bias in animal behaviour research: Can we believe what we score, if we score what we believe? *Animal Behaviour*, *90*, 273–280.

Vilette, C., Bonnell, T., Henzi, P., & Barrett, L. (2020). Comparing dominance hierarchy methods using a data-splitting approach with real-world data. *Behavioral Ecology*, *31*(6), 1379–1390.

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., … Vázquez-Baeza, Y. (2020). Scipy 1.0: Fundamental algorithms for scientific computing in python. *Nature Methods*, *17*(3), 261–272.

Wooddell, L. J., Kaburu, S. S., Suomi, S. J., & Dettmer, A. M. (2017). Elo-rating for tracking rank fluctuations after demographic changes involving semi-free–ranging rhesus macaques (*Macaca mulatta*). *Journal of the American Association for Laboratory Animal Science*, *56*(3), 260–268.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

Figure S1. Bottom-up ranking of 'Langley2018' dataset using ADAGIO method.

Figure S2. (A) Illustration of the AniDomNet hierarchy rank evolution over time for one simulated case ($\alpha = 0.2$). A short interval of disruption of the dominance ranks can be observed. (B) Illustration of the Optimized Elo rating scores for the same dataset as comparison to the AniDomNet. While there is no rank changes between Animal 3 and 5, inconsistency in the ranking of Animal 4 can be observed around $t = 50$.

Figure S3. Recovery rate of the true dominance hierarchy under varying group sizes and known dyad percentages.

Table S1. Benchmark results: AniDomNet parametrization.

How to cite this article: Ipek, N., Tuyttens, F. A. M., De Baets, B., & Verwaeren, J. (2025). AniDomNet: A sequential pairwise model for inferring dynamic animal dominance hierarchies. *Methods in Ecology and Evolution*, *16*, 2145–2163. https://doi.org/10.1111/2041-210X.70118